Large-area mountain pine beetle infestations: spatial data representation and accuracy

Trisalyn Nelson[1], Barry Boots[2], and Michael A. Wulder[3]

Mountain Pine Beetle Initiative
Working Paper 2006–02

Natural Resources Canada | Ressources naturelles Canada

Canada

This publication is funded by the Government of Canada through the Mountain Pine Beetle Initiative, a program administered by Natural Resources Canada, Canadian Forest Service.

Natural Resources Canada, Canadian Forest Service,
Pacific Forestry Centre, 506 West Burnside Road, Victoria, BC V8Z 1M5
(250) 363-0600 • www.pfc.cfs.nrcan.gc.ca

2007-00169

Canada

Mountain Pine Beetle Initiative
Canadian Forest Service

Large-area mountain pine beetle infestations: spatial data representation and accuracy

Trisalyn Nelson[1], Barry Boots[2], and Michael A. Wulder[3]

Mountain Pine Beetle Initiative
Working Paper 2006–02

[1]University of Victoria, Department of Geography,
PO Box 3050, Victoria, British Columbia, V8W 3P5, Canada

[2]Wilfrid Laurier University, Department of Geography and Environmental Studies,
75 University Ave West, Waterloo, Ontario, N2L 3C5, Canada

[3]Canadian Forest Service,Pacific Forestry Centre, Natural Resources Canada,
506 West Burnside, Victoria, British Columbia, V8Z 1M5, Canada

This working paper was previously published in the *Forestry Chronicle*

Corresponding Author: Trisalyn Nelson (address above) email: trisalyn@uvic.ca fax: 250-721-6216; phone: 250-472-5620

Natural Resources Canada
Canadian Forest Service
Pacific Forestry Centre
506 West Burnside Road
Victoria, British Columbia V8Z 1M5
Canada

©Her Majesty the Queen in Right of Canada 2006
Printed in Canada

Abstract

Point data generated from helicopter surveys are used to determine the location and magnitude of mountain pine beetle infestations. Although collected for tactical planning, these data also provide a rich source of information for scientific investigations. To facilitate spatial research, it is important to consider how to best represent spatially explicit mountain pine beetle infestation data. This paper focuses on the spatial representation of point-based aerial helicopter surveys, which can be difficult to represent due to issues associated with large data quantities and data uncertainty. In this paper, the benefit of using a kernel density estimator to convert point data to a continuous raster surface is demonstrated. Field data are used to assess the accuracy of the point-based aerial helicopter survey data and the kernel density estimator is extended to incorporate data uncertainty. Although the accuracy of point-based aerial surveys is high, with 92.6% of points differing by no more than ±10 trees, there is a general tendency to overestimate infestation magnitude. The method developed for incorporating uncertainty into the kernel density estimator reduces overestimation and improves the correspondence between estimated infestation intensities and field data values.

Keywords: Mountain pine beetle, data representation, visualization, kernel density estimators, uncertainty

Résumé

Des points de données tirées de sondage par hélicoptère sont utilisés pour déterminer la localisation et l'importance des infestations de dendroctone du pin. Même si elles ont été recueillies pour des raisons de planification tactiques, ces données constituent également une importante source d'information pour les études scientifiques. Afin de faciliter la localisation spatiale, il est important d'étudier comment on peut représenter le mieux les données explicites en terme de localisation des infestations de dendroctone du pin. Cet article porte sur la représentation spatiale des points tirés des sondages par hélicoptère, ce qui peut être difficile à réaliser compte tenu des questions entourant les grandes quantités de données et l'incertitude qu'elles comportent. Dans le cas présent, l'avantage de l'utilisation d'un estimateur de noyau de densité pour la conversion des points de données en surface continue est démontré. Des données de terrain sont utilisées pour évaluer la précision des points de données du sondage aérien par hélicoptère et l'estimateur de noyau de densité est conçu pour comprendre l'incertitude entourant les données. Même si la précision des points de données par sondage aérien est élevée, 92 % des points ne différant par pas plus de ± 10 arbres, on retrouve une tendance générale de surestimer l'étendue de l'Infestation. La méthode élaborée pour incorporer l'incertitude au sein de l'estimateur de noyau de densité réduit la surestimation et améliore la correspondance entre les intensités d'infestation estimées et les valeurs en provenance des données du terrain.

Mots clés : dendroctone du pin, représentation des données, visualisation, estimateurs de noyau de densité, incertitude

Table of Contents

Abstract / Résumé 3
Introduction 7
Study Area and Data Description 9
Methods 10
  Data Representation Via Kernel Density Estimation 10
  Aerial Survey Accuracy Assessment 11
  Extending Kernel Density Estimation to Include Uncertainty 11
Results and discussion 13
  Data Representation Via Kernel Density Estimation 13
  Aerial Survey Accuracy Assessment 14
  Extending Kernel Density Estimation to Include Uncertainty 16
Conclusions 18
Acknowledgements 18
References 19

Introduction

Mountain pine beetle populations in British Columbia, Canada, are increasing exponentially. Warmer temperatures combined with an over-abundance of mature lodgepole pine have led to an epidemic that is much larger than any previously recorded. The current mountain pine beetle infestation has affected 7 million ha of British Columbia's forest (Westfall 2005), and projections indicate that 80% of the lodgepole pine may be affected unless a major weather-stopping event occurs (Eng et al. 2004).

As the epidemic grows, so does recognition that forest management requires a new understanding of mountain pine beetle dynamics. Fundamental research questions surrounding landscape-level, spatial dynamics of the mountain pine beetle await exploration. These questions include research that would improve knowledge of spatial processes such as how mountain pine beetle disperse and select hosts. Such insights would benefit modelling efforts and decision-making. As well, the current knowledge regarding the mountain pine beetle is typically based on stand- or finer-scale studies, and little evidence is available to indicate if processes operate similarly at the landscape level.

The coupling of improvements in spatial data acquisition with intensive forest monitoring efforts has led to the availability of large-area, spatially explicit data on the impacts of the mountain pine beetle. Used to target mitigation and other activities, spatial point data on infestations are a valuable asset for forest planners. A spin-off benefit is that spatially explicit data sets provide scientists with a new mechanism for investigating large-area spatial mountain pine beetle dynamics.

In order for empirical mountain pine beetle data to be used for scientific purposes, an optimal approach for data representation is required. While large-area spatial mountain pine beetle data sets are collected using various approaches (Wulder et al. 2004), the focus of this paper is point-based, aerial survey data. Such data are generated via helicopter surveys during which the spatial location of an infestation cluster is marked using a GPS and the number of infested trees is estimated. Unless specifically noted, the mountain pine beetle aerial survey data referenced throughout this manuscript are GPS point data collected through helicopter surveys, and differ from polygon format aerial survey data collected from fixed-wing aircraft and used to support broader-scale strategic planning.

There are several issues to consider when representing point-based mountain pine beetle data collected via aerial surveying. The first issue, which is a conceptual consideration for all spatial data representation, is the nature of mountain pine beetle processes. Although endemic populations mark the landscape with periodic and spatially discrete pine mortality, the mountain pine beetle is ubiquitous in pine forests (Carroll and Safranyik 2004). During an epidemic, pine mortality becomes spatially continuous, with infested trees occurring throughout the forest and discrete boundaries becoming difficult to delineate. Particularly when data reflect epidemic mountain pine beetle conditions, there is a conceptual benefit to using a spatially continuous data representation.

A second factor to consider in representing aerial survey point data on beetle populations is that the points correspond to areas, but the size of the corresponding areas varies and is typically not recorded. A spatial representation that links points to areas may be advantageous; however, standard area conversions are difficult due to variability in the size of areas represented by each point.

The large number of points collected during aerial surveying is a third consideration in data representation. Point-based mountain pine beetle data sets have many points, and the associated number of infested trees ranges from one to 400. As a result, when using point representations, visualization of the spatial variation in infestation intensity is obscured by symbol

overlap. Standard cartographic methods, such as proportional symbols or colours, become insufficient to allow for meaningful interpretation of data.

Data uncertainty is a fourth consideration when representing aerial survey data on mountain pine beetle infestations. All aerial surveys of mountain pine beetle infestations use indicators of pine mortality, mainly changes in crown foliage colour, to monitor mountain pine beetle activity. When a pine tree is attacked by the mountain pine beetle, usually in mid-summer, crown foliage changes successively from green to yellow, to brown, to red, and eventually needles drop off the tree, leaving the grey stem and branches (Safranyik et al. 2002). Although the first visually detectable change to foliage colour usually occurs in the spring following attack, hot and/or dry summer conditions may cause foliage to begin changing in the fall (Safranyik et al. 2002). Typically, one year after attack, crown foliage is yellow green or yellow brown (Safranyik et al. 2002). By two years after the attack, the foliage often becomes red and needles fall off in the third year. The rate of foliage change varies substantially. For instance, crown foliage may become red in a single year, and trees may retain red and brown needles for as long as three years (Safranyik et al. 2002). Most often, by the time foliage has turned red and attacks are clearly visible, mountain pine beetle have emerged and moved to a new host (Safranyik et al. 2002). Uncertainty resulting from the variable rate of foliage discolouration impacts scientific investigation. As the number of mountain pine beetle infestation data points increase, so do the impacts of uncertainty, and errors can compound when multi-temporal data are investigated. Although it is always beneficial to consider data uncertainty, when mountain pine beetle data are used for large-area spatial and spatial-temporal investigations, representations that deal with error are necessary to ensure the integrity of analysis results.

The goal of this paper is to describe an appropriate method for representing point-based, landscape-scale data on mountain pine beetle when: 1) data sets are large, and 2) data are uncertain. To meet this goal, three objectives were the focus:

1. A kernel density estimator is presented as an effective approach for representing large, point-based, mountain pine beetle data as continuous surfaces (grids);
2. The nature of uncertainty in point-based, aerial survey mountain pine beetle data is explored through comparisons with field data;
3. The kernel density estimator is extended to generate a representation of the mountain pine beetle data that incorporates data uncertainty.

Study Area and Data Description

The Morice Timber Supply Area (TSA), in western British Columbia, Canada (Figure 1), is currently experiencing epidemic numbers of mountain pine beetle. Covering an area of approximately 1.5 million ha, Morice includes five biogeoclimatic zones: Sub-Boreal Spruce (59%), Engelman Spruce–Subalpine Fir (26%), Alpine Tundra (11%), Coastal Western Hemlock (3%), and Mountain Hemlock (1%). Morice is bordered by Tweedsmuir Provincial Park in the south; topography is gentle to the north and east, becoming more mountainous in the southwest. There are three major rivers in the area (the Bulkley, the Morice, and the Nadina) and two large lakes (Babine and Ootsa).

Managers of the Morice TSA monitor mountain pine beetle infestations using point-based, GPS aerial surveys. During aerial surveys, clusters of visually infested trees are identified, and a GPS is used to map cluster centroids as points. For each cluster, the number of infested trees is estimated and the infesting insect species recorded. This data set is unique in that the number of infested trees, rather than an infestation severity class, is assigned to each point. The maximum area represented by a point is 0.031 km^2, equivalent to a circle with a radius of 100 m. Although many points represent smaller areas, variations in circle radii are not recorded. From 1999 to 2002, 5 682 field data were collected for aerial survey points. Field surveys verified the cause of lodgepole pine mortality. If trees were killed by mountain pine beetle, the number of infested trees and the timing of attack was recorded.

In this work, the aerial survey accuracy was assessed using all field data. However, development and verification of representation methods focus on data collected in 2002. The 2002 aerial survey has 8 777 points, and field data were collected for 2 820 of locations.

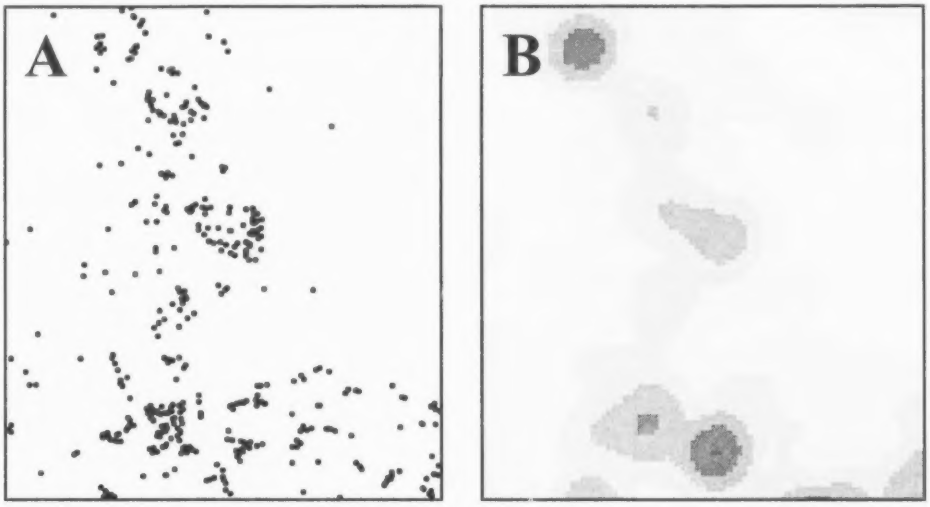
Figure 1. Morice TSA in Western British Columbia, Canada, is centered on Houston (54° 24' N, 126° 38' W)

Methods

Data Representation Via Kernel Density Estimation

Kernel density estimators are a powerful and flexible method for generating a spatially continuous representation of point data (Silverman 1986; Gatrell 1994; Bailey and Gatrell 1995). Instead of representing clusters of infested trees as discrete events (points), kernel density estimators can be used to create a continuous surface that displays the intensity of infested trees over the study area (Figure 2). A continuous representation of the mountain pine beetle data has conceptual benefits, improves visualization, and overcomes issues of point-area associations.

Figure 2. Kernel density estimators are used to convert points (A) to continuous surfaces (grid) (B). Darker coloured cells have higher infestation intensities.



Conceptually, the intensity $\lambda(z)$ at a particular location **z** in a study area *A* can be estimated by the naïve kernel density estimator:

$$\hat{\lambda}(z) = \frac{\text{the number of events in a disk centred on } z}{\text{area of the disk}} \qquad [1]$$

A more precise estimate, $\hat{\lambda}_\tau(\mathbf{z})$ is defined by:

$$\hat{\lambda}_\tau(z) = \frac{1}{p_\tau(z)}\left\{\sum_{i=1}^{n}\frac{1}{\tau^2}k\left(\frac{(z-z_i)}{\tau}\right)y_i\right\} \; z \in A \qquad [2]$$

where **z** and *A* are defined as above, τ is the radius of a disk centred on **z**, $k(\,)$ is the kernel or a probability density function that is symmetric around about the origin, $\mathbf{z}_i$ ($i = 1, \ldots, n$) are locations of *n* observed events, and y_i is the attribute value at $\mathbf{z}_i$. The term $p_\tau(z) = \int_A k[(z-\boldsymbol{u})/\tau]d\boldsymbol{u}$ is an edge correction equivalent to the volume under the scaled kernel centred on **z** which lies inside of *A* (Diggle 1985).

There are three issues that are commonly considered when working with kernel density estimators: the type of kernel $k(\ )$, the size of disk radius τ, and edge effects. The kernel determines how events within the disk radius will be weighted. For example, in this research a kernel with a normal distribution function, which weights events in terms of their distance from $\mathbf{z}$, was used. Although kernel definition may be theoretically important, it does not have a large impact on kernel output.

Kernel density estimators are much more sensitive to τ, which controls data smoothing. As τ increases, so does the amount of data smoothing (Kelsall and Diggle 1995); if τ is too large, data variability will be lost; if too small, data trends will not be visible. Ideally, kernels need to be calculated for several values of τ and kernel outputs compared. In this study, a 2-km disk radius was chosen to optimize tradeoffs between detail and representations of infestation trends. The 2-km disk radius is also large enough to be relatively robust with respect to errors in point locations, and is reflective of stand-scale research, which indicates that a high proportion of local beetle dispersal occurs within 2 km of emergence (Safranyik et al. 1992).

Kernel density estimators may also be impacted by edge effects. In this study, edge effects do not have a large impact, as the study area is large relative to τ, so the edge-correction term in Equation 2 was not implemented. An additional issue, which arises in software used to implement kernel density estimators, is the definition of raster cell size. Here, a 200m x 200m grid cell is used, which is approximately equal to the maximum area represented by an aerial survey point.

In this study, to compare the kernel density method to other standard representations of point data, the aerial survey data were displayed using: points, proportional symbols, and proportional colours.

Aerial Survey Accuracy Assessment

The availability of field data to assess the accuracy of point-based aerial surveys was capitalized upon. As field data does not include information on the spatial error of points, this assessment focuses on the the number of infested trees recorded for each point location. However, Morice TSA data collectors and users have suggested that the spatial error of aerial survey points is approximately ±25 m (Nelson et al. 2004).

Data collected in the field were manually partitioned based on the timing of the mountain pine beetle attack. This enabled us to link the aerial survey data, indicative of visibly detectable attack, to infestations that were recorded as occurring one year previously. Green attack is a recent infestation that has not yet led to a visible response in lodgepole pine foliage. Green attack, also recorded in the field, is not used for accuracy assessment, as it is not captured by aerial surveys. In the remainder of this paper, reference to field data indicates only data for infestations that were recorded as occurring one year prior to the aerial survey.

An error matrix was used to assess the correspondence of the number of infested trees collected via field and aerial surveys. For low numbers of infested trees (1 to 15), each value is considered a unique category in the error matrix. Higher numbers of infested trees, with fewer occurrences, are grouped into the following classes: 16 to 20; 21 to 25; 26 to 50, and; 51 to 400.

Extending Kernel Density Estimation to Include Uncertainty

By extending the kernel density estimator to incorporate uncertainty, data error can be considered when spatially representing point data on mountain pine beetle infestations. Generally, the approach to incorporating uncertainty in kernel surfaces is to generate a number of possible realizations of the aerial survey point data, produce a kernel density surface for each, and compute the average infestation intensity for all surfaces (Figure 3).

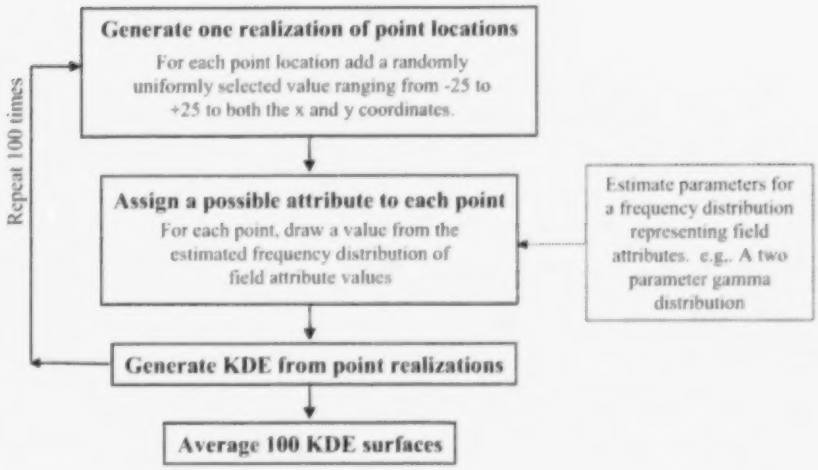

Generate one realization of point locations
For each point location add a randomly uniformly selected value ranging from -25 to +25 to both the x and y coordinates.
Repeat 100 times
Assign a possible attribute to each point
For each point, draw a value from the estimated frequency distribution of field attribute values
Estimate parameters for a frequency distribution representing field attributes. e.g., A two parameter gamma distribution
Generate KDE from point realizations
Average 100 KDE surfaces

Figure 3.

Summary of method for incorporating the spatial and attribute uncertainty of points when generating kernel density-estimated surfaces.

The distribution of GPS error in either the *x*- or *y*-axis is typically near normal (Leva et al. 1996). Therefore, spatial error was simulated by randomly selecting values from a normal distribution with a mean of zero and standard deviation of 1. Drawn values were scaled between ±25 and added to both the *x* and *y* locations of each point. For each point, this procedure was carried out 100 times, allowing 100 realizations of aerial survey point locations.

In this analysis, field values are considered to represent true infestation levels. The frequency distribution of field values can therefore be used as the basis for simulating realizations of the number of infested trees identified during aerial surveys. Since there are many low values in the field data and the associated frequency distribution is skewed right, the values can be modelled by a gamma distribution. However, the gamma is a positively valued, continuous distribution, whereas the number of infested trees is discrete and includes zeros (reflecting aerial survey errors). Thus, two corrections have to be made in order to use the model gamma distribution in simulating the number of infested trees. First, zero values are randomly assigned to the same proportion of aerial survey points as field data sites. Then, discrete values are generated for the remaining aerial survey points by binned values drawn from the gamma distribution (Yang 1994).

Although these solutions may lack elegance, the alternative is to generate a distribution based on the observed data, which would be "chunky" and requires the assumption that the observed data represent the entire error population, rather than a sample of possible errors. Furthermore, this paper focuses on developing an approach to incorporating data uncertainty when data are represented continuously using kernel density estimation. Although this method requires a model of data error, the error model is not the focus.

A kernel density surface was estimated for each of 100 realizations of the point data incorporating both spatial and attribute value uncertainty. These 100 surfaces were then averaged to produce a single continuous representation of infestation intensity that incorporated uncertainty.

To verify the effectiveness of the correction method, field data were compared to estimates generated using the basic kernel density estimator and the kernel density estimator modified to incorporate uncertainty. Field data from 2002 were gridded using a cell size of 200m x 200m to allow spatially explicit comparisons with kernel density-estimated values. Frequency histograms of the differences between field and kernel-estimated values were used to verify the effectiveness of the correction.

Results and discussion

Data Representation Via Kernel Density Estimation

In Figure 4, a basic point representation of the aerial survey data (4a) is compared with proportional symbol mapping (4b), proportional grey-tone mapping (4c), and kernel density estimation (4d). Visual comparisons indicate that kernel density estimators lead to a more effective representation of point data. By generating a continuous representation of the data, kernel density estimators remove point overlap and spread attribute values over areas. Improvements in data representation are both conceptual and applied, with kernel density estimators improving visualization and communication of infestation conditions.

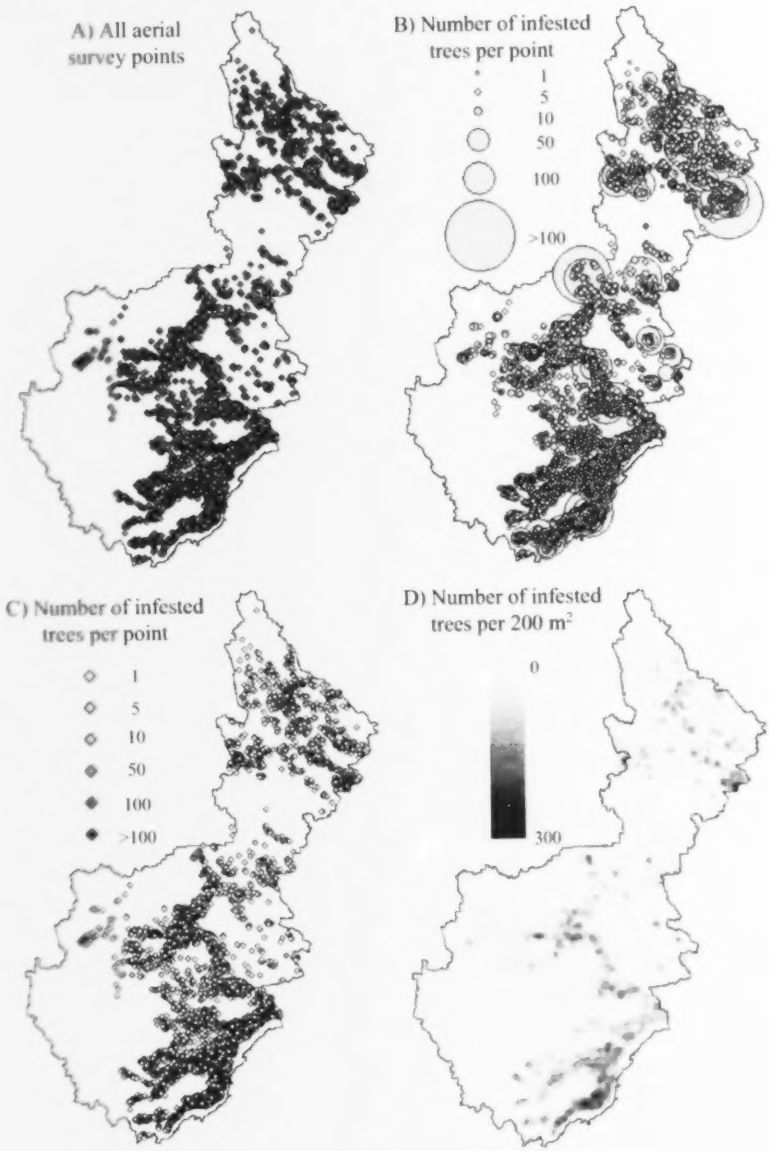


Figure 4. Comparison of data visualization techniques. A) Aerial survey points with no enhancements; B) Aerial survey point attributes represented as proportional symbols; C) Aerial survey point attributes represented as proportional grey tones; D) Aerial survey point attributes represented using a kernel density estimator with a 2-km disk radius, normal kernel shape, and 200-m cell size.

Aerial Survey Accuracy Assessment

Overall, only 17.4% of points had the same number of infested trees identified in both the field and aerial surveys, and commission and omission errors are high. However, it should be recalled that the aerial surveys are intended as estimates and are not meant to be exact counts of infested trees. For the purpose of management, estimates provide sufficient detail and accuracy for planning and mitigation. Thus, it is worth noting that 34.6% of aerial survey points have an error ranging between ±1 tree; 50.5% have errors between ±2 trees; 80.5% have errors ranging between ±5 trees, and; 92.6% have errors of ±10 trees (Table 1).

Table 1. Percentage of aerial survey points with the correct number of infested trees. Accuracy is determined through comparison with field data and calculated for several error ranges. 17.4% of aerial survey points have the same number of infested trees in both field and aerial surveys. However, 92.6% of aerial survey points have numbers of infested trees that deviate only ± 10 trees relative to field attributes.

Error range (number of trees)	Aerial survey points having the number of infested trees within the error range
0	17.4%
± 1	34.6%
± 2	50.5%
± 5	80.5%
± 10	92.6%

The investigation into the nature of uncertainty in the number of infested trees is summarized in the error matrix shown in Table 2 (next page). This error matrix is not the standard square form, as there are no values of zero when the number of infested trees is determined during aerial surveys. Associations between the number of infested trees and errors of omission or commission are weak. The dominant trend noted in the error matrix is the prevalence for aerial surveys to overestimate the number of infested trees. While the reason for overestimation cannot be determined from the existing data, a possible explanation is the variable rate for foliage discolouration. It is likely that aerial surveyors are identifying additional trees that were infested less than one year previously and have a fast foliage response, as well as trees attacked more than one year previously that have a slow foliage response. Other causes of overestimation could include misidentification of the cause of tree mortality and operational issues such as surveyor inexperience.

These results do not provide a true appraisal of omission error. As the location of field surveys are based on the presence of infested trees detected during aerial surveys, field data provide no information on any infested trees missed entirely during aerial surveying. Factors that impact omission errors and generate underestimates include: variable rates of crown foliage discolouration; survey sample design, and; operational issues such as aircraft speed, flying height, and weather conditions.

Table 2. Error matrix comparing the number of infested trees detected during all years of aerial and field surveys. Columns represent the number of attacked trees detected in the field; rows represent the number of attacked trees detected during aerial surveys.

Aerial/Field	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16-20	21-25	26-50	51-400	n	Commission
1	69	123	6	3	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	204	39.7
2	110	76	157	34	15	8	4	1	3	0	1	0	0	0	0	0	1	0	0	0	410	61.7
3	388	236	233	223	69	35	26	15	8	5	1	4	2	2	1	1	3	2	4	2	1260	82.3
4	198	69	107	106	109	41	25	6	9	8	3	4	1	2	1	2	1	2	1	0	695	84.3
5	227	67	67	99	97	118	46	37	21	13	11	6	2	8	2	4	9	3	1	1	839	85.9
6	107	25	34	60	33	39	52	24	19	11	13	10	5	2	3	3	3	2	1	0	446	88.3
7	36	10	12	20	25	25	31	23	15	11	6	5	3	1	2	2	1	0	1	1	230	90.0
8	52	17	18	17	23	23	21	23	35	16	17	9	4	5	5	3	12	3	2	0	305	88.5
9	12	3	2	2	3	6	4	6	3	10	2	1	0	2	0	0	1	0	1	0	58	82.8
10	89	13	21	24	11	17	17	20	17	10	39	17	13	17	20	9	16	8	6	6	390	90.0
11	2	1	4	1	0	1	0	2	1	2	2	4	0	0	0	1	1	0	0	0	22	81.8
12	26	5	7	3	4	7	4	10	6	0	10	3	11	6	4	5	9	4	4	0	128	91.4
13	6	2	0	1	0	0	0	0	0	2	0	1	0	3	0	1	1	0	0	0	17	82.4
14	6	3	0	0	1	1	0	0	1	0	0	0	0	1	2	0	0	0	1	0	16	87.5
15	44	2	5	2	4	6	8	4	6	6	8	3	8	2	6	9	30	20	14	2	189	95.2
16-20	36	3	3	4	3	1	2	0	2	4	6	4	1	6	4	7	17	5	1	3	112	84.8
21-25	20	3	3	2	0	0	1	0	0	3	1	0	1	0	0	3	13	5	7	3	65	92.3
26-50	34	5	0	4	1	2	2	1	2	2	3	0	0	0	0	2	14	6	16	5	99	83.8
51-400	11	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	4	3	5	2	27	92.6
n	1473	664	679	605	400	330	243	172	149	103	123	72	51	57	50	52	136	63	65	25	5512	
Omission	0.00	81.5	76.8	63.1	72.8	64.2	78.6	86.6	76.5	90.3	68.3	94.4	78.4	94.7	96.0	82.7	87.5	92.1	75.4	92.0		

Table 2. Error matrix comparing the number of infested trees detected during all years of aerial and field surveys. Columns represent the number of attacked trees detected in the field; rows represent the number of attacked trees detected during aerial surveys.

Aerial/Field	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16-20	21-25	26-50	51-400	n	Commission
1	69	123	6	3	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	204	39.7
2	110	76	157	34	15	8	4	1	3	0	1	0	0	0	0	0	1	0	0	0	410	61.7
3	388	236	233	223	69	35	26	15	8	5	1	4	2	2	1	1	3	2	4	2	1260	82.3
4	198	69	107	106	109	41	25	6	9	8	3	4	1	2	1	2	1	2	1	0	695	84.3
5	227	67	67	99	97	118	46	37	21	13	11	6	2	8	2	4	9	3	1	1	839	85.9
6	107	25	34	60	33	39	52	24	19	11	13	10	5	2	3	3	3	2	1	0	446	88.3
7	36	10	12	20	25	25	31	23	15	11	6	5	3	1	2	2	1	0	1	1	230	90.0
8	52	17	18	17	23	23	21	23	35	16	17	9	4	5	5	3	12	3	2	0	305	88.5
9	12	3	2	2	3	6	4	6	3	10	2	1	0	2	0	0	1	0	1	0	58	82.8
10	89	13	21	24	11	17	17	20	17	10	39	17	13	17	20	9	16	8	6	6	390	90.0
11	2	1	4	1	0	1	0	2	1	2	2	4	0	0	0	1	1	0	0	0	22	81.8
12	26	5	7	3	4	7	4	10	6	0	10	3	11	6	4	5	9	4	4	0	128	91.4
13	6	2	0	1	0	0	0	0	0	2	0	1	0	3	0	1	1	0	0	0	17	82.4
14	6	3	0	0	1	1	0	0	1	0	0	0	0	1	2	0	0	0	1	0	16	87.5
15	44	2	5	2	4	6	8	4	6	6	8	3	8	2	6	9	30	20	14	2	189	95.2
16-20	36	3	3	4	3	1	2	0	2	4	6	4	1	6	4	7	17	5	1	3	112	84.8
21-25	20	3	3	2	0	0	1	0	0	3	1	0	1	0	0	3	13	5	7	3	65	92.3
26-50	34	5	0	4	1	2	2	1	2	2	3	0	0	0	0	2	14	6	16	5	99	83.8
51-400	11	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	4	3	5	2	27	92.6
n	1473	664	679	605	400	330	243	172	149	103	123	72	51	57	50	52	136	63	65	25	5512	
Omission	0.00	81.5	76.8	63.1	72.8	64.2	78.6	86.6	76.5	90.3	68.3	94.4	78.4	94.7	96.0	82.7	87.5	92.1	75.4	92.0		

Several approaches to large-area aerial surveying of mountain pine beetle infestations use infestation classes, rather than the number of infested trees (Wulder et al. 2004). Common infestation classes are: 1 to 5; 6 to 10, and; greater than 10. Therefore, to extend the understanding of the information provided by aerial surveying, it is helpful to consider the nature of data uncertainty when infestation-intensity classes are employed. An error matrix associated with the above infestation classes is presented in Table 3 and indicates that errors of commission and omission increase for larger infestation classes. These results indicate that, to maximize accuracy in the estimated number of infested trees, there may be benefit to aerial survey designs that favour many survey locations (points) with small infestation classes. Also, regardless of the infestation class size, commission errors are more frequent then omission errors. These results are consistent with error analysis for the non-classed numbers of infested trees.

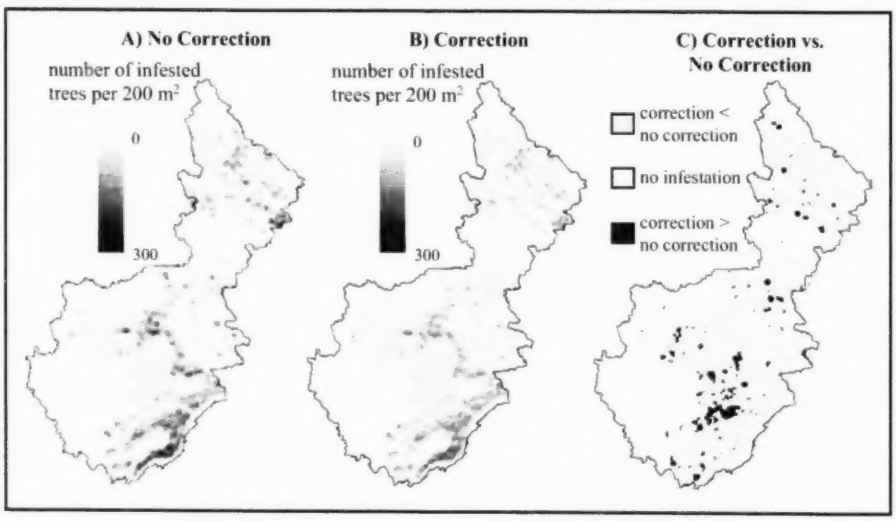
Table 3. Error matrix comparing classes of the number of infested trees detected during all years of aerial and field surveys. Columns represent classes of the number of attacked trees detected in the field; rows represent classes of the number of attacked trees detected during aerial surveys.

Aerial/Field	0	1-5	6-10	>10	n	Commission
1-5	992	2100	244	72	3408	38.38
6-10	296	483	445	205	1429	68.86
>10	185	95	94	294	668	55.99
n	1473	2678	783	571	5505	
Omission	0	21.58	43.17	48.51		

Extending Kernel Density Estimation to Include Uncertainty

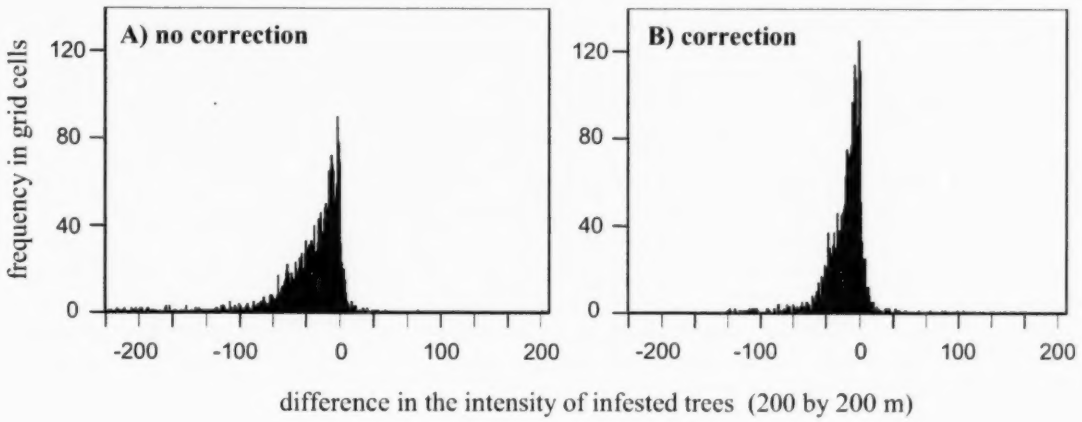
Representations of mountain pine beetle point data using a standard kernel density estimation approach and a modification that considered data uncertainty are shown in Figures 5a and 5b, respectively. Locations on the figures with zero values are considered to have no infestation. The differences in corrected and uncorrected values are mapped in Figure 5c. Generally, the correction leads to lower estimated values, which is encouraging, given the tendency to overestimate the number of infested trees during aerial surveying. Locations where the correction produces a lower estimate of infestation intensity occur primarily at the edges of the spatial extent of the epidemic.

Figure 5. Kernel density estimated surfaces for 2002: A) Kernel density estimate with no correction; B) Kernel density estimate with correction; C) Differences between corrected and uncorrected kernel density estimated surfaces.



The frequency distributions of differences between field values and kernel densit-estimated values generated with and without consideration of uncertainty are shown in Figure 6. When interpreting these frequency histograms, it is interesting to note that kernel density estimators smooth data. As a result, an underlying effect of kernel density estimation is a reduction in the magnitude of values. The drawback to such an approach is that actual extreme infestation intensities could be artificially reduced. However, given that most errors in point aerial survey values occur as overestimates, the basic kernel density estimator is likely to implicitly lead to a data representation that is more reflective of actual conditions. Results indicate that this is so, as kernel density estimators that incorporate uncertainty lead to better correspondence between estimates of infestation intensity and field data. The distribution of errors associated with the uncorrected scenario is skewed further to the left, showing a greater tendency for kernel density-estimated values to be higher than field values. As well, the frequency distribution of errors associated with kernel density estimates that incorporate uncertainty has a smaller range and more values at or near zero.

Figure 6. Difference between gridded field and kernel-density estimated values. A) Kernel density values are not corrected; B) Kernel density values are corrected.



Conclusions

Kernel density estimators are a simple and effective tool for representing infestation data and facilitate exploration and mapping of spatial variation in infestation magnitude. Conceptually, there are benefits to representing point data on mountain pine beetle infestations as continuous surfaces, implemented as a raster grid, using kernel density estimation. Improvements to representation enhance the use of data for large-area spatial investigations of mountain pine beetle dynamics. Kernel density estimators also improve data visualization, allowing more effective communication with decision-makers and improved public information dissemination. Through kernel density estimation, the study area is partitioned into grid cells, and multi-temporal investigations are enabled as grid cells can be compared over many time periods.

With 92.6% of aerial survey points having errors of ±10 trees, the quality of data is appropriate for management and monitoring efforts. However, these rich data may also be used for science, and, in the case of large-area spatial or spatial-temporal analysis, small errors can have cumulative impacts and need to be accounted for. Point-based helicopter aerial surveys tend to generate overestimates of infestation levels. As a smoothing technique, the basic kernel density estimator helps deal with this. However, kernel density estimators modified to incorporate data uncertainty further improve representation of infestation intensities.

The method demonstrated for incorporating uncertainty when undertaking kernel density estimation may be applied to other applications when data sets are large, and uncertainty and error can be modelled. In future work, kernel density-estimated surfaces, generated from point-based GPS data on mountain pine beetle infestations, can be used to generate a new spatial understanding of landscape-scale beetle dynamics.

Acknowledgements

This project was funded by the Government of Canada through the Mountain Pine Beetle Initiative (MPBI), a six-year, $40-million Program administered by Natural Resources Canada, Canadian Forest Service. Publication does not necessarily signify that the contents of this report reflect the views or policies of Natural Resources Canada, Canadian Forest Service. Additional information on the MPBI is available at: http://mpb.cfs.nrcan.gc.ca/. The authors would also like to thank the Morice Timber Supply Area for providing data, and anonymous reviewers for input that improved this manuscript.

References

Bailey, T.; Gatrell, A. 1995. Interactive Spatial Data Analysis. Essex: Longman Group Limited. 413 pp.

Diggle, P. 1985. A kernel method for smoothing point process data. Applied Statistics 34:138–147.

Carroll A.L.; Safranyik, L. 2004. The bionomics of the mountain pine beetle in lodgepole pine: establishing a context. Pages 1–32 in T.L. Shore, J.E. Brooks, and J.E. Stone, editors. Mountain Pine Beetle Symposium: Challenges and Solutions, October 30–31, 2003. Kelowna, British Columbia, Canada. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia, Information Report BC-X-399. 298 pp.

Eng, M.; Fall, A.; Hughes, J.; Shore, T.; Riel, B.; Hall, P. 2003. Provincial Level Projection of the Current Mountain Pine Beetle Outbreak: An Overview of the Model (BCMPB) and Draft Results of Year 1 of the Project, Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, BC. Mountain Pine Beetle Initiative Working Paper 2004-01. 39 pp.

Gatrell, A. 1994. Density estimation and the visualization of point patterns. pages 65–75 in H. Hernshaw and D. Uwin, Visualization in Geographic Information Systems. Chichester: John Wiley & Sons.

Kelsall, J.; Diggle, P. 1995. Non-parametric estimation of spatial variation in relative risk. Statistics in Medicine 14:2335–2342.

Leva, J.; Uijt de Haag, M.; Dyke, K. 1996. Performance of standalone GPS. In: E.D. Kapplan, Understanding GPS: Principles and Applications. Boston: Artech House Publishers. 279 pp..

Nelson, T.; Boots, B.; Wulder, M. 2004. Beetle infestations to characterize pattern, risk, and spread at the landscape level. Pages 164–173 in T.L. Shore, J.E. Brooks, and J.E. Stone, editors. Mountain Pine Beetle Symposium: Challenges and Solutions, October 30–31, 2003, Kelowna, British Columbia, Canada. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia, Information Report BC-X-399. 298 pp.

Safranyik, L.; Linton, D.; Silversides, R.; McMullen, L. 1992. Dispersal of released mountain pine beetles under the canopy of a mature lodgepole pine stand. Journal of Applied Entomology 113:441–450.

Safranyik, L.; Shrimpton, D.; Whitney, H. 2002. Management of Lodgepole Pine to Reduce Losses from the Mountain Pine Beetle: Second Edition. Victoria: Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre. 23 pp.

Silverman, B. 1986. Density Estimation for Statistics and Data Analysis. New York: Chapman Hall. 175 pp.

Westfall, J. 2005. Summary of Forest Health Conditions in British Columbia. British Columbia Ministry of Forests, Forest Practices Branch. 43 pp.

Wulder, M.A.; Dymond, C.; Erikson, R.D. 2004. Detection and monitoring of the mountain pine beetle. Natural Resources Canada, Canadian Forest Service, Pacific Forestry Centre, Victoria, British Columbia, Information Report BC-X-398. 24 pp.

Yang, Z. 1994. Maximum likelihood phylogentic estimation from DNA sequences with variable rates over sites: approximation methods. Journal of Molecular Evolution 39:306–314.

This publication is funded by the Government of Canada through the Mountain Pine Beetle Initiative, a program administered by Natural Resources Canada, Canadian Forest Service (web site: mpb.cfs.nrcan.gc.ca).

Contact:

For more information on the Canadian Forest Service, visit our web site at:
www.nrcan.gc.ca/cfs-scf

or contact the Pacific Forestry Centre
506 West Burnside Road
Victoria, BC V8Z 1M5
Tel: (250) 363-0600 Fax: (250) 363-0775
www.pfc.cfs.nrcan.gc.ca

To order publications on-line, visit the Canadian Forest Service Bookstore at:
bookstore.cfs.nrcan.gc.ca